# Representing Music Using MIMO Models for Genre Clustering

Bram Geelen Bart De Moor
KU Leuven, Department of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics
bram.geelen@esat.kuleuven.be; bart.demoor@esat.kuleuven.be

## 1 Overview

We present a method to analyse musical works, by representing melodies as time series of pitch class activations. For any musical work - represented as audio or as a digital score - we can create a *beat-aligned chromagram*, which describes the activity of the twelve *pitch classes* (A, A#, ..., G#) at every beat. Previous work has shown that simple statistical descriptors of such a chromagram can be used to perform adequate genre classification [3], yet these descriptors do not incorporate the pitch class transitions - which are essential to the *harmonical progression* of a song - into the song representation. If we construct an AR model of these transitions, we exactly create a model of how the melody and harmonies progress throughout the song. With the (flattened) matrices of the AR model of every song, we can perform a simple genre clustering using off-the-shelf methods.

## 2 Clustering beat-aligned chromagrams

To construct the desired state representation from an audio source, we must first create the *constant-Q spectrogram* of the analysed song [1]. This is very similar to a short-time Fourier Transform, in that N narrow-band filters are convolved with the input signal, yet with *harmonically* spaced frequencies instead of linearly spaced. A chromagram aggregates the activities of the result into bins of each of the 12 musical *pitch classes*. Next, we locate the *beats* in the song, with the beat tracking algorithm proposed by Ellis [2]. Then we simply average the activity of of the pitch classes within each beat, to receive the beat-aligned chromagram.
We can also create a beat-aligned chromagram from digital music score representations such as `.midi`-files, by first identifying the beat of the song (often this is inherently part of the score), and subsequently aggregating per beat the activity of the 12 pitch classes throughout the song.
We then wish to represent the harmonic progression of the song as a vector, to ease the further work of clustering the corpus. We do this by looking at the beat-aligned chromagram as a time series of vector states $X_t$ (with every vector representing a single beat), and constructing a simple rank-one MIMO AR model $M$;

$$X_{t+1} = M \cdot X_t$$

We can then flatten the matrix $M$ to receive the vector representation we want. Note that we can increase the accuracy of this model by increasing its rank, yet this will result in a worse song representation; a least-squares solution for $M$ could differ heavily between similar songs, especially with high correlation in the input space. To alleviate this problem, we can simply *correlate* the pitch class activities from one state to the next. This means we *don't* create a higher-order AR model, but only analyse how every pitch class activation is correlated with every pitch class in the next beat (or $k$ beats *further*). Thus, we create a correlation matrix $M$ from the beat-aligned chromagram $X$ (of shape $(T,12)$) as follows;

$$M_{ij} = corr(X_{1:T-k,i}, X_{k:T,j})$$

In this presentation, we will show how to interpret both the beat-aligned chromagrams as well as the correlation matrices, and demonstrate how they differ between genres. We will also show how the representations perform in benchmark genre classification datasets, when combined with off-the-shelf classifying models.

## Acknowledgements

This work was supported in part by the KU Leuven Research Fund (projects C16/15/059, C32/16/013, C24/18/022), in part by the Industrial Research Fund (Fellowship 13-0260) and several Leuven Research and Development bilateral industrial projects, in part by Flemish Government Agencies: FWO (EOS project 30468160 (SeLMA), SBO project I013218N (Alamire), PhD grants (SB/1SA1319N, SB/1S93918, SB/151622)), EWI (Flanders AI Impulse Program), VLAIO (City of Things (COT.2018.018), industrial projects (HBC.2018.0405), and PhD grants: Baekeland mandate (HBC.20192204) and Innovation mandate (HBC.2019.2209)), and in part by the European Commission (EU H2020-SC1-2016-2017 Grant Agreement 727721: MIDAS).

## References

[1] Judith C Brown. Calculation of a constant q spectral transform. *The Journal of the Acoustical Society of America*, 89(1):425–434, 1991.

[2] Daniel PW Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[3] Alexander Schindler and Andreas Rauber. Capturing the temporal domain in echonest features for improved classification effectiveness. In *International Workshop on Adaptive Multimedia Retrieval*, pages 214–227. Springer, 2012.